

Deployment of 5G Networking Infrastructure with Machine Type Communication Considerations

Xiangxiang Xu[†], Walid Saad*, Xiujun Zhang[†], Limin Xiao[†] and Shidong Zhou[†]

[†] State Key Laboratory on Microwave and Digital Communications,

Tsinghua National Laboratory for Information Science and Technology,

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

* Wireless@VT, Bradley Department of Electrical and Computer Engineering, Blacksburg, VA, USA.

Emails: xuxxmail@163.com, walids@vt.edu, zhangxiujun@tsinghua.edu.cn, xiaolm@mail.tsinghua.edu.cn, zhousd@mail.tsinghua.edu.cn

Abstract—Designing optimal strategies to deploy small cell stations is crucial to meet the quality-of-service requirements in next-generation cellular networks with constrained deployment costs. In this paper, a general deployment framework is proposed to jointly optimize the locations of backhaul aggregate nodes, small base stations, machine aggregators, and multi-hop wireless backhaul links to accommodate both human-type and machine-type communications. The goal is to provide deployment solutions with best coverage performance under cost constraints. The formulated problem is shown to be a multi-objective integer programming for which it is challenging to obtain the optimal solutions. To solve the problem, a heuristic algorithm is proposed by combining Lagrangian relaxation, the weighted sum method, the ϵ -constraint method and tabu search to obtain both the solutions and bounds, for the objective function. Simulation results show that the proposed framework can provide solutions with better performance compared with conventional deployment models in scenarios where available fiber connections are scarce. Furthermore, the gap between obtained solutions and the lower bounds is quite tight.

I. INTRODUCTION

In next-generation cellular networks, low power small base stations (SBSs) will be densely deployed to boost the coverage and capacity [1, 2]. Furthermore, billions of machines will coexist with conventional human-type communication (HTC) [3, 4]. Thus, next-generation cellular systems must integrate heterogeneous base stations and heterogeneous types of communications. All these components should be considered in the network planning period to obtain a network deployment strategy with optimal operating performance. Indeed, one major challenge for next-generation cellular networks is to design flexible, scalable and low cost deployment strategies [2] that capture the network's heterogeneity.

For SBSs, backhaul links are required to connect them to the core network. The performance of an SBS also depends on the capacity of the backhaul link. Thus, backhaul links and SBSs should be considered jointly during network planning [2] to guarantee the system performance. Conventional works on base station (BS) deployment have focused mainly on scenarios

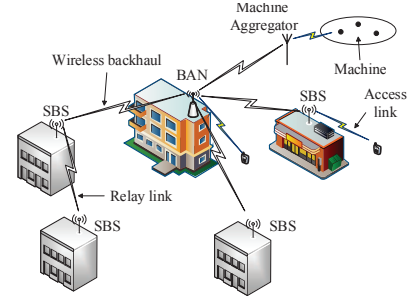


Fig. 1 System model.

in which fiber connection backhaul links are available for all candidate locations [5–9]. In other words, only sites with existing fiber connections are regarded as potential sites. However, as discussed in [2], the availability of existing fiber connections might be scarce which would hamper the flexibility and performance of the network. Also, deploying new fiber connections for dense SBSs is unacceptable from the perspective of cost. To overcome this problem, the works in [2, 10] have shown that heterogeneous backhaul, wired and wireless, would be a viable solution for ultra dense small cell network. The authors in [11] studied the deployment of wireless backhaul links for SBSs. However, this work only studied the placement of wireless backhaul nodes which provide backhaul links to SBSs directly, through one hop, while SBSs deployment has been neglected.

Although interesting deployment problems have been studied, existing works [5–9, 11] have only focused on fibre backhaul links and HTC. Such works can not be extended to solve the deployment problem of the next-generation networks since the impact of MTC and the designing of backhaul network have been neglected. In particular, heterogeneous communications, HTC and machine-type communication (MTC), and heterogeneous backhaul links, wired and wireless, should be considered jointly in the planning period. In addition, multi-hop wireless backhaul links should be allowed to improve the network performance when the available fiber links are scarce. To our best knowledge, the general framework of the deployment of next-generation network while taking into account jointly HTC, MTC, wired backhaul and wireless backhaul, has never been addressed yet.

The main contribution of this paper is to propose a practical general deployment framework which incorporates HTC, MTC, available wired backhaul links and multi-hop wireless backhaul links for next-generation cellular networks. Three types of stations, backhaul aggregate nodes (BANs), small base stations (SBSs) and machine aggregators (MAs), are deployed to get the network with optimal performance. In particular, MAs are

This work was supported by National Basic Research Program of China (Grant No. 2012CB316002), National Natural Science Foundation of China (Grant No. 61201192), Science Fund for Creative Research Groups of NSFC (61321061), National S&T Major Project (Grant No. 2015ZX03001016-002), National High Technology Research and Development Program of China (863 Program) (Grant No. 2014AA01A704), Tsinghua University Initiative Scientific Research (2015Z02-3), International Science and Technology Cooperation Program (Grant No. 2012DFG12010), Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (Grant No. 2012D02), Tsinghua-Qualcomm Joint Research Program, the Tsinghua-Intel International S&T Cooperation Program (ICRI-MNC), and the U.S. Office of Naval Research (ONR) under Grant N00014-15-1-2709.

responsible for covering MTC and backhaul links of an SBS can be provided either by some BAN or by other SBSs. We formulate the deployment problem as a multi-objective integer programming which is NP-hard. To obtain the entire Pareto optimal solution set, we propose an tabu search based adaptive algorithm which can also provide lower bounds for the problem and the bounds can be used to check the quality of the obtained solutions. Simulation results show that our framework can get good solutions in scenarios where available fiber connections are scarce. Furthermore, the ratio of the objective value of the obtained solutions to the lower bounds can be as tight as 1.99.

The rest of the paper is organized as follows. Section II presents the system and problem formulation. The proposed deployment algorithm is introduced and discussed in Section III. Section IV analyzes the simulation results and conclusions are drawn in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Our model is shown in Fig. 1. The backhaul aggregate nodes (BANs) are placed at sites with available fiber connections to connect SBSs to the core network. Meanwhile, the deployed SBSs with backhaul can act as relays and connect other SBSs to BANs. Thus, for an SBS, the backhaul link can be provided either by a BAN or by other SBSs. In addition, we assume that BANs have access ability and can cover a certain area.

For MTC, a clustered infrastructure [4, 12] is adopted in which the machines first transmit data to a machine aggregator (MA) which will then send the data to a BAN as shown in Fig. 1. In the system, we mainly consider static machines whose locations and data rates are assumed to be known. Unlike HTC, for MTC, uplink transmission is our focus in the deployment period.

For HTC, both access links and backhaul links are assumed to work in the orthogonal millimeter-wave (mmW) bands. For MTC, a conventional, sub-6 GHz band is used for access since MTCs require long range, low rate coverage, which is in contrast to mmW which is a high bandwidth, short-range solution. For mmW propagation, the pathloss model in [13] is used:

$$Loss(d) = 20\log_{10}\left(\frac{4\pi d_0}{\lambda}\right) + 10\bar{n}\log_{10}\left(\frac{d}{d_0}\right) + \chi, \quad (1)$$

where d_0 is a reference distance and λ is the wavelength of the carrier. \bar{n} denotes the path loss exponent and d is transmission distance. χ is the shadowing component which is a zero mean Gaussian variable with a standard deviation σ_s in dB. The value \bar{n} and σ_s depend on the transmission conditions, i.e., line of sight (LOS) or non line of sight (NLOS), access link or backhaul link. The LOS probability derived in [14] is utilized in our work, $p_{LOS}(d) = e^{-\beta d}$, where d denotes transmission distance and β is related with transmission environment.

A. Problem formulation

Consider an area \mathcal{A} in which a number of BANs, SBSs and MAs must be deployed to serve both HTC and MTC. In this area, the potential sites set is given by $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Set $\mathcal{F}_1 = \{1, \dots, F_1\}$ indicates sites with available fiber connection and denotes the candidate locations for BANs. Set $\mathcal{F}_2 = \{1, \dots, F_2\}$ and $\mathcal{F}_3 = \{1, \dots, F_3\}$ denote the candidate locations for SBSs and MAs, respectively.

We define the deployment variables $z_k \in \{0, 1\}$, $y_i \in \{0, 1\}$ and $w_j \in \{0, 1\}$ for BANs, SBSs and MAs, respectively, to indicate the station is deployed ($= 1$) or not ($= 0$). For HTC, in order to indicate the area coverage, we partition the area \mathcal{A} equally into a set of subareas which are elements of the set $\mathcal{S} = \{1, \dots, S\}$. For each subarea $s \in \mathcal{S}$, the center point is chosen to indicate its coverage. Variables x_{ks} and x_{is} are defined as follows to indicate the coverage of subarea s by BAN k and SBS i , respectively.

$$x_{ks} \ (x_{is}) = \begin{cases} 1 & \text{if subarea } s \text{ is covered by BAN } k \text{ (SBS } i). \\ 0 & \text{otherwise.} \end{cases}$$

As previously mentioned, we assume that SBSs can act as relays. To bound the transmission latency, it is reasonable to restrict the number of hops from an SBS to a BAN. Here, N is used to denote the maximum number of relays for a connection from an SBS to a BAN. Then, the maximum number of hops from an SBS to a BAN is $N + 1$. For the multi-level network, we define the following variables to indicate the connections among BSs.

$$x_{ki} \ (x_{pi}) = \begin{cases} 1 & \text{if BAN } k \text{ (SBS } p) \text{ provides backhaul to SC } i. \\ 0 & \text{otherwise.} \end{cases}$$

Here, we let $x_{ii} = 0$ for all $i \in \mathcal{F}_2$. It needs to be clarified that the level of our network infrastructure is not fixed. The maximum level is $N + 1$ and some SBSs can connect to BANs through just one or two hops. For subarea s , we define the *downlink* routing variables $x_{ki}^s \in \{0, 1\}$ ($x_{pi}^s \in \{0, 1\}$) to indicate whether the data for s is routed from BAN k (SBS p) to SBS i or not.

For MTC, we let $\mathcal{M} = \{1, \dots, M\}$ be the set of the machines that need to be covered by MAs. Each machine $m \in \mathcal{M}$ has a data rate d_m . We define variables $x_{jm} \in \{0, 1\}$ and $x_{kj} \in \{0, 1\}$ to indicate the coverage of machine $m \in \mathcal{M}$ by MA $j \in \mathcal{F}_3$ and the availability of a link between BAN k and MA j , respectively.

Given the area \mathcal{A} , our purpose is to find an optimal deployment solution which indicates the locations of BANs, SBSs, MAs and their connections. From an operator's perspective, we want to maximize the coverage with a minimum deployment cost. In order to examine the relationship between deployment cost and performance deeply, unlike the previous studies [5–9], we consider multiple objectives simultaneously and give an optimal deployment solution set from which the proper strategy can be chosen. The three objective functions are defined as follows.

$$f_1(\mathbf{z}, \mathbf{y}, \mathbf{w}) = \sum_{k \in \mathcal{F}_1} z_k c_k + \sum_{i \in \mathcal{F}_2} y_i c_i + \sum_{j \in \mathcal{F}_3} w_j c_j, \quad (2)$$

$$f_2(\mathbf{x}) = S - \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{F}_1} x_{ks} - \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{F}_2} x_{is}, \quad (3)$$

$$f_3(\mathbf{x}) = M - \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{F}_3} x_{jm}, \quad (4)$$

where $f_1(\mathbf{z}, \mathbf{y}, \mathbf{w})$ represents deployment cost and $f_2(\mathbf{x})$, $f_3(\mathbf{x})$ indicate the number of uncovered subareas and machines. c_k , c_i and c_j denote deployment cost for BAN k , SC i and MA j , respectively. Vector $\mathbf{x} = [x_{ki}, x_{ks}, x_{is}, x_{ki}^s, x_{ip}^s, x_{jm}]$, $\mathbf{y} = [y_i]$, $\mathbf{z} = [z_k]$ and $\mathbf{w} = [w_j]$ are the binary variables.

In the considered area \mathcal{A} , mobile users are assumed to be distributed according to a homogeneous Poisson Point Process (PPP) of density λ_u . A subarea is said to be covered if the outage

probability is less than a predefined threshold, p_{oa} . Then, the deployment problem can be formulated as a multi-level multi-objective integer programming problem:

$$\begin{aligned}
& \text{(P)} \quad \min_{\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}} [f_1(\mathbf{y}, \mathbf{z}, \mathbf{w}), f_2(\mathbf{x}), f_3(\mathbf{x})], \\
& \text{s.t.} \quad x_{ks} \leq z_k \quad s \in \mathcal{S}, k \in \mathcal{F}_1, \quad (5) \\
& \quad x_{is} \leq y_i \quad s \in \mathcal{S}, i \in \mathcal{F}_2, \quad (6) \\
& \quad x_{ki} \leq y_i \quad k \in \mathcal{F}_1, i \in \mathcal{F}_2, \quad (7) \\
& \quad x_{ki} \leq z_k \quad k \in \mathcal{F}_1, i \in \mathcal{F}_2, \quad (8) \\
& \quad x_{pi} \leq y_p \quad i, p \in \mathcal{F}_2, \quad (9) \\
& \quad x_{pi} \leq y_i \quad i, p \in \mathcal{F}_2, \quad (10) \\
& \quad x_{ki}^s \leq x_{ki} \quad s \in \mathcal{S}, k \in \mathcal{F}_1, i \in \mathcal{F}_2, \quad (11) \\
& \quad x_{ip}^s \leq x_{ip} \quad s \in \mathcal{S}, i, p \in \mathcal{F}_2, \quad (12) \\
& \quad \sum_{k \in \mathcal{F}_1} x_{ks} + \sum_{i \in \mathcal{F}_2} x_{is} \leq 1 \quad s \in \mathcal{S}, \quad (13) \\
& \quad \sum_{k \in \mathcal{F}_1} x_{ks} p(\gamma_{ks} \leq \gamma_t) + \sum_{i \in \mathcal{F}_2} x_{is} p(\gamma_{is} \leq \gamma_t) \leq p_{oa}, \quad (14) \\
& \quad \sum_{i \in \mathcal{F}_2} x_{ki} + \sum_{j \in \mathcal{F}_3} x_{kj} \leq N_b \quad k \in \mathcal{F}_1, \quad (15) \\
& \quad \sum_{k \in \mathcal{F}_1} x_{ki} + \sum_{p \in \mathcal{F}_2} x_{pi} = y_i \quad i \in \mathcal{F}_2, \quad (16) \\
& \quad \sum_{k \in \mathcal{F}_1} \sum_{i \in \mathcal{F}_2} x_{ki}^s = \sum_{i \in \mathcal{F}_2} x_{is} \quad s \in \mathcal{S}, \quad (17) \\
& \quad x_{is} + \sum_{p \in \mathcal{F}_2} x_{ip}^s = \sum_{k \in \mathcal{F}_1} x_{ki}^s + \sum_{p \in \mathcal{F}_2} x_{pi}^s, \quad (18) \\
& \quad \sum_{p \in \mathcal{F}_2} \sum_{i \in \mathcal{F}_2} x_{pi}^s + \sum_{k \in \mathcal{F}_1} \sum_{i \in \mathcal{F}_2} x_{ki}^s \leq N + 1 \quad s \in \mathcal{S}, \quad (19) \\
& \quad \sum_{p \in \mathcal{F}_2} x_{pi} p(r_i > C_{pi}) + \sum_{k \in \mathcal{F}_1} x_{ki} p(r_i > C_{ki}) \leq p_{ob}, \quad (20) \\
& \quad x_{kj} \leq z_k \quad k \in \mathcal{F}_1, j \in \mathcal{F}_3, \quad (21) \\
& \quad \sum_{j \in \mathcal{F}_3} x_{jm} \leq 1 \quad m \in \mathcal{M}, \quad (22) \\
& \quad x_{jm} d_{jm} \leq D_{jt} \quad j \in \mathcal{F}_3, m \in \mathcal{M}, \quad (23) \\
& \quad \tau_j \leq \tau_{th} \quad j \in \mathcal{F}_3, \quad (24) \\
& \quad \sum_{k \in \mathcal{F}_1} x_{kj} = w_j \quad j \in \mathcal{F}_3, \quad (25) \\
& \quad \sum_{m \in \mathcal{M}} x_{jm} d_m \delta \leq C_{kj} \quad k \in \mathcal{F}_1, j \in \mathcal{F}_3. \quad (26)
\end{aligned}$$

Here, γ_{ks} (γ_{is}) denotes the received signal to noise ratio (SNR) at the center point of subarea s from BAN k (SBS i). γ_t is an outage threshold for user access. N_b is the maximum number of nodes that a BAN can serve. The random variable r_i indicates the total data request from the covered area of SBS i . C_{pi} (C_{ki}) is the capacity of the link from SBS p (BAN k) to BS i . p_{ob} is the maximum allowable probability with which the limited backhaul capacity induces large transmission latency. d_{jm} is the distance from MA j to machine m and D_{jt} indicates the coverage range of MA j . τ_{th} denotes the transmission latency threshold of MTC. $0 < \delta \leq 1$ represents the data compression ratio.

In problem (P), constraints (5), (6), (7), (8), (9), (10) and (21) mean that only deployed BSs can provide connections to

users, other BSs and machines. (11) and (12) guarantee that a connection is required to route data among SBSs. (13) and (14) ensure that each subarea can be covered at most by one BS with a lower outage probability than the predefined threshold. Since millimeter wave band networks are noise limited [10, 13], SNR is used to indicate the quality of service. Constraint (15) indicates the backhaul capability of a BAN. Constraint (16) ensures that a *deployed* SBS ($y_i = 1$) must have backhaul links either through connecting to BANs or other SBSs. Also, we do not consider the case in which one SBS has access to multiple BANs or SBSs simultaneously. For subarea s , constraint (17) implies that the data for s must be transmitted from a BAN to some SBS once it is covered by an SBS ($\sum_{i \in \mathcal{F}_2} x_{is} = 1$). (18) means that the

data for subarea s can be transmitted to SBS i when subarea s is covered by SBS i ($x_{is} = 1$) or SBS i relays data to some other SBS ($\sum_{p \in \mathcal{F}_2} x_{ip}^s = 1$). Constraint (19) is the constraint on the number of hops. (20) characterizes the wireless backhaul capacity limits. Here, we mainly focus on the limited wireless backhaul capacity and assume that access capacity is sufficient.

Constraint (22) means that the machine can be covered by at most one MA and (23) is the transmission distance constraint for machines. (24) ensures the mean uplink transmission delay of machines covered by MA j is less than the predefined threshold. Constraint (25) requires that, for deployed MAs, some deployed BANs should provide wireless backhaul links. (26) implies that the backhaul capacity of an MA is sufficient to satisfy the aggregate machine demand. Here, we assume that the data rate of an machine is quite low and it can be satisfied by the connected MA. In problem (P), the hop limits and constraints (20) on the large latency probability together guarantee that a long path does not exist in the deployed networks.

Based on channel model (1), constraint (14) can be transformed into the following constraint on the transmission distance.

$$\sum_{i \in \mathcal{F}_2} x_{is} (d_{is} - D_{it}) + \sum_{k \in \mathcal{F}_1} x_{ks} (d_{ks} - D_{kt}) \leq 0. \quad (27)$$

d_{is} (d_{ks}) is defined as the distance from SBS i (BAN k) to the center point of subarea s . D_{it} and D_{kt} denote the coverage range of SBS i and BAN k , respectively.

Since users are distributed according to a PPP, (20) can be transformed further to the following constraints on the number of covered subareas given the users' rate demand distribution.

$$\sum_{s \in \mathcal{S}} x_{is} + \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{F}_2} x_{ip}^s \leq N_i(\mathbf{x}) \quad i \in \mathcal{F}_2. \quad (28)$$

Here, $N_i(\mathbf{x}) = \sum_{k \in \mathcal{F}_1} x_{ki} N_{ki} + \sum_{p \in \mathcal{F}_2} x_{pi} N_{pi}$. N_{ki} (N_{pi}) denotes the number of subareas that SBS i can cover when connected to BAN k (SBS p). In [3], the author proposed a framework to compute the mean transmission latency τ_j for MA j in MTC which is a function of the number of machines and resources. Then, given available resources and latency requirement τ_{th} , the method proposed in [3] can be modified to obtain the following constraint on the total number of machines that an MA can cover.

$$\sum_{m \in \mathcal{M}} x_{jm} \leq N_j w_j \quad j \in \mathcal{F}_3. \quad (29)$$

Problem (P) is a *multi-objective multi-level capacitated facility location problem (FLP)*. This problem is an extension of multi-level FLP which is known to be NP-hard [15]. Here, Pareto optimality is considered for the multi-objective programming.

Definition 1: For a minimization programming with objective function $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_N(\mathbf{x})]$. A solution \mathbf{x}^* is *efficient* if there does not exist other solution \mathbf{x} such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$, $\forall i$ and $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$, $\exists i$. The point $\mathbf{f}(\mathbf{x}^*)$ in the objective space is called *Pareto optimal* or *nondominated* if \mathbf{x}^* is efficient.

Hereinafter, we use efficient and Pareto optimal interchangeably. Our purpose is to find the optimal solution set and guide the practical deployment of BANs, SBSs and MAs.

III. PROPOSED ALGORITHM

To solve problem (P), we first combine the weighted sum method with the ϵ -constraint method [16] together to transform it to a single objective optimization problem. In the weighted sum method, we assume that the network planner has a parameter θ that captures the tradeoff between the HTC coverage and MTC coverage. Then, the coverage objective function can be expressed as $f_c(\mathbf{x}) = f_2(\mathbf{x}) + \theta f_3(\mathbf{x})$. Then, through ϵ -constraint method, the deployment cost $f_1(\mathbf{y}, \mathbf{z}, \mathbf{w})$ is taken as a constraint. Given θ and ϵ , the problem is transformed to the following problem.

$$\begin{aligned} (\mathbf{P}_\epsilon) \quad & \min f_c(\mathbf{x}), \\ \text{s.t.} \quad & f_1(\mathbf{z}, \mathbf{y}, \mathbf{w}) \leq \epsilon, \\ & (5-13), (15-19), (21-23), (25-29). \end{aligned} \quad (30)$$

The number of binary variables in problem (\mathbf{P}_ϵ) is considerably large. Conventional algorithm like branch-and-bound algorithm and linear relaxation can not address the problem efficiently. Here, we adopt a scheme that integrates Lagrangian relaxation and tabu search together to solve (\mathbf{P}_ϵ) .

A. Lagrangian relaxation

We note that the backhaul capacity constraint (28) makes the problem challenging to handle. So we introduce Lagrangian multipliers $\lambda = \{\lambda_i \geq 0, i \in \mathcal{F}_2\}$ and relax constraint (28). Then, the relaxed problem $(\mathbf{P}_\lambda^\epsilon)$ is obtained.

$$\begin{aligned} (\mathbf{P}_\lambda^\epsilon) \quad & \min S + \theta M - \sum_{k \in \mathcal{F}_1} m_k + \sum_{i \in \mathcal{F}_2} n_i - \theta \sum_{j \in \mathcal{F}_3} \sum_{m \in \mathcal{M}} x_{jm}, \\ \text{s.t.} \quad & (5-13), (15-19), (21-23), (25-27), (29), (30), \end{aligned}$$

where

$$\begin{aligned} m_k &= \sum_{s \in \mathcal{S}} x_{ks} + \sum_{i \in \mathcal{F}_1} \lambda_i N_{ki} x_{ki}, \\ n_i &= \sum_{s \in \mathcal{S}} (\lambda_i - 1) x_{is} + \lambda_i \sum_{p \in \mathcal{F}_2} \left(\sum_{s \in \mathcal{S}} x_{ip}^s - x_{pi} N_{pi} \right). \end{aligned}$$

Here, we define a path as a sequence of nodes that originates from some BAN and ends at some SBS. Path $P = \{k, i_1, \dots, i_q\}$ indicates that BAN k (hop 0) serves SBS i_1 and SBS i_n (hop n) relays data to i_{n+1} with $1 \leq n \leq (q-1)$. For SBS i , P_i and h_i are defined to denote its path and the hop number, respectively. The following parameters are defined to facilitate description of proposed algorithms.

Algorithm 1 Local search algorithm for problem $(\mathbf{P}_\lambda^\epsilon)$ with fixed BSs.

Input: $[\mathbf{z}, \mathbf{y}, \mathbf{w}]$, θ , λ_i , N_b , N , N_{ki} , N_{ip} , d_m , D_{kt} , D_{it} , D_{jt} ($k \in \mathcal{F}_1, i, p \in \mathcal{F}_2, j \in \mathcal{F}_3, m \in \mathcal{M}$).

- 1: initialize $\mathbf{x} = \mathbf{0}$, $\mathcal{K} = \{k | z_k = 1\}$, $\mathcal{I} = \{i | y_i = 1\}$, $\mathcal{W} = \{j | w_j = 1\}$, $P_i^n = 0$, $\mathcal{P}_{i,b} = \emptyset$, $\mathcal{P}_{i,a} = \emptyset$, $L_{P_i} = 0$, $r_i = 0$, $\mathcal{I}_o = \emptyset$.
- 2: for s , $\mathcal{K}_s = \{k | z_k = 1, d_{ks} \leq D_{kt}\}$, $k = \arg \min_{k' \in \mathcal{K}_s} d_{k's}$, $x_{ks} = 1$.
- 3: **while** ($\mathcal{W} \neq \emptyset$)
- 4: for $j \in \mathcal{W}$, $k \in \mathcal{K}$, denote the maximum coverage of j when $x_{kj} = 1$ as Cov_{kj} .
- 5: $\{j_0, k_0\} = \arg \max_{j \in \mathcal{W}, k \in \mathcal{K}} \text{Cov}_{kj}$, $x_{k_0 j_0} = 1$, update $x_{j_0 m}$, remove MA j_0 from \mathcal{W} .
- 6: if $\sum_{j \in \mathcal{F}_3} x_{k_0 j} = N_b$, remove BAN k_0 from \mathcal{K} .
- 7: **end while**
- 8: compute the objective value of $(\mathbf{P}_\lambda^\epsilon)$ and denote as V .
- 9: **while** ($\mathcal{I} \neq \emptyset$)
- 10: for $i \in \mathcal{I}$, $k \in \mathcal{K}$, denote the maximum coverage of i when $x_{ki} = 1$ as Cov_{ki} , denote $\Delta V_{ki} = -\text{Cov}_{ki}$.
- 11: for $i \in \mathcal{I}$, $p \in \mathcal{I}_o$, if $L_p < N + 1$, denote $\Delta V_{ip} = \{\text{Objective value change when } P_p^{h_p-1} = i\}$ and $\Delta V_{pi} = \{\text{Objective value change when } P_p^{h_p+1} = i\}$.
- 12: $\{i_0\} = \arg \min_{i \in \mathcal{I}} \{\Delta V_{ki}, \Delta V_{pi}, \Delta V_{ip}, k \in \mathcal{K}, p \in \mathcal{I}_o\}$.
- 13: if $\Delta V_{k_0 i_0} = \Delta V = \min_{i \in \mathcal{I}} \{\Delta V_{ki}, \Delta V_{pi}, \Delta V_{ip}, k \in \mathcal{K}, p \in \mathcal{I}_o\}$, $x_{k_0 i_0} = 1$, update path P_{i_0} and $L_{P_{i_0}}$. if $\sum_{i \in \mathcal{I}_o} x_{k_0 i} + \sum_{j \in \mathcal{F}_3} x_{k_0 j} = N_b$, remove BAN k_0 from \mathcal{K} .
- 14: if $\Delta V_{p_0 i_0} = \Delta V$, insert i_0 between p_0 and $P_{p_0}^{h_{p_0}-1}$; if $\Delta V_{i_0 p_0} = \Delta V$, insert i_0 between p_0 and $P_{p_0}^{h_{p_0}+1}$. $P_{i_0} = P_{p_0}$, update path P_{i_0} .
- 15: remove i_0 from \mathcal{I} and $\mathcal{I}_o = \mathcal{I}_o \cup \{i_0\}$, $V = V + \Delta V$.
- 16: **end while**

Output: \mathbf{x} , $V(\mathbf{z}, \mathbf{y}, \mathbf{w}) = V$

$$\begin{aligned} P_i^n &= \{i' | P_{i'} = P_i, h_{i'} = n\}, \\ \mathcal{P}_{i,b} &= \{i' | i' \in \mathcal{F}_2, P_{i'} = P_i, h_{i'} < h_i\}, \\ \mathcal{P}_{i,a} &= \{i' | i' \in \mathcal{F}_2, P_{i'} = P_i, h_{i'} > h_i\}, \\ L_{P_i} &= \max\{h_{i'} | i' \in \mathcal{F}_2, P_{i'} = P_i\}, \\ r_i &= \sum_{s \in \mathcal{S}} x_{is}. \end{aligned}$$

Then n_i can be replaced by

$$n_i' = \left(\sum_{p \in \mathcal{P}_{i,b}} \lambda_p + \lambda_i - 1 \right) r_i - \lambda_i N_{P_i^{(h_i-1)} i}.$$

Since $(\mathbf{P}_\lambda^\epsilon)$ is a relaxation of (\mathbf{P}_ϵ) , a lower bound for (\mathbf{P}_ϵ) can be obtained through solving $(\mathbf{P}_\lambda^\epsilon)$. However, problem $(\mathbf{P}_\lambda^\epsilon)$ is still no less complex than the NP-Complete knapsack problem since the benefits of BSs, m_k and n_i , depend on each other closely. If we relax constraint (19) further, the obtained optimal value of $(\mathbf{P}_\lambda^\epsilon)$ would degrade significantly. Here, we propose a local search algorithm to obtain the best connection variables \mathbf{x} and optimal objective value $V(\mathbf{z}, \mathbf{y}, \mathbf{w})$ given deployment variables $[\mathbf{z}, \mathbf{y}, \mathbf{w}]$ which is described in Algorithm 1.

In line 11 in the Algorithm 1, for SBS i , we provide the following procedure to compute the ΔV_{ip} .

Step 1: $\Delta V_{ip} = 0$. For SBS $p' \in \mathcal{P}_{p,a}$, if $\sum_{i' \in \mathcal{P}_{p',b}} \lambda_{i'} + \lambda_i > 1$,

$$\Delta V_{ip} = \Delta V_{ip} - \left(\sum_{i' \in \mathcal{P}_{p',b}} \lambda_{i'} - 1 \right) r_{p'} \text{ and } r_{p'} = 0.$$

Step 2: For those p' with $r_{p'} = 0$, $x_{p's} = 0$ for $s \in \mathcal{S}$.

Step 3: If $\sum_{i' \in \mathcal{P}_{p,b}} \lambda_{i'} + \lambda_i > 1$, $r_i = 0$. If $\sum_{i' \in \mathcal{P}_{p,b}} \lambda_{i'} + \lambda_i \leq 1$, based on the updated connection variables in step 2, determine r_i of SBS i when it is connected to $P_p^{h_p-1}$. $\Delta V_{ip} = \Delta V_{ip} + (\sum_{i' \in \mathcal{P}_{p,b}} \lambda_{i'} + \lambda_i - 1)r_i$.

Step 4: If $P_p^{h_p-1}$ is BAN k_0 , $\Delta V_{ip} = \Delta V_{ip} + \lambda_p(N_{k_0p} - N_{ip}) - \lambda_i N_{k_0i}$.

Step 5: If $P_p^{h_p-1}$ is SBS i_0 , $\Delta V_{ip} = \Delta V_{ip} + \lambda_p(N_{i_0p} - N_{ip}) - \lambda_i N_{i_0i}$.

ΔV_{pi} can be calculated in a similar way. Based on Algorithm 1, next, we propose a tabu search based algorithm to solve (P_λ^ϵ) .

B. Tabu search

Tabu search is a local search scheme that can avoid being trapped in local optimal solutions through a short term memory called tabu list. Here, we propose a tabu search algorithm to solve (P_λ^ϵ) and obtain Pareto optimal solution set for (P).

Here, we define three local moves to search the neighborhood $\mathcal{N}(z, y, w)$ of a solution $[z, y, w]$: 1) *Open move*: deploy a station at an empty site; 2) *Close move*: remove a deployed station; 3) *Swap move*: remove a deployed station and deploy a new station. It needs to be noted that the objective values of solutions in $\mathcal{N}(z, y, w)$ can not be obtained from the objective value of $[z, y, w]$. Optimal variables x given deployment variables can only be got through solving the problem.

To solve (P_λ^ϵ) first, we propose a two-level tabu search algorithm described in Algorithm 2 to get the optimal solution $[z, y, w, x]$ and optimal value $V_B = V(z, y, w)$. In Algorithm 2, restart diversification scheme is used: re-deploy N_{div} rarely deployed stations when there is no non-tabu solutions in the neighborhood. Meanwhile, the two-level structure acts as an intensification scheme to search the solution space of SBSs and MAs thoroughly. In Algorithm 2, the Algorithm 1 is used in each iteration, line 9 and 14, to calculate the objective value $V(z', y', w')$ for each solution $(z', y', w') \in \mathcal{N}(z, y, w)$.

Given the output of the Algorithm 2, $[z, y, w]$, a two-level search algorithm, as included in Algorithm 3 from line 9 to line 19, is executed to obtain Pareto optimal solution set for problem (P). Unlike Algorithm 2, multiple Pareto optimal solutions rather than one optimal solution can be got. In the algorithm, a new parameter $\Delta\epsilon$ is introduced to control the intensification scheme. In each iteration, only solution space with cost in $[\epsilon - \Delta\epsilon, \epsilon]$ are searched. It needs to be noted that in the search of solutions for problem (P), objective functions $f_1(z, y, w)$ and $f_c(x)$ are used to determine whether a solution is efficient or not.

C. Update of ϵ

In order to get the entire optimal solution set, ϵ needs to be decreased gradually. All nondominated solutions can be obtained through the update of ϵ [16]. Given ϵ_t in iteration t , optimal solution set \mathcal{P}_{ϵ_t} can be obtained through the proposed algorithm. Then, ϵ_t can be updated adaptively as follows:

$$\epsilon_{t+1} = \min \left(\min_{[z', y', w'] \in \mathcal{P}_{\epsilon_t}} f_1(z', y', w'), \epsilon_t \right) - \Delta c.$$

When $\mathcal{P}_{\epsilon_t} = \emptyset$, we update the cost constraint based on ϵ_t . Δc is a small positive number.

The algorithm to solve (P) is summarized as Algorithm 3.

Algorithm 2 Algorithm for problem (P_λ^ϵ) .

Input: $N_{t1}, N_{t2}, N_{swap}, N_{div}$, tabu list \mathcal{T}_1 and \mathcal{T}_2 , ϵ , θ , λ_i , N_b , N , N_{ki} , N_{ip} , d_m , D_{kt} , D_{it} , D_{jt} ($k \in \mathcal{F}_1, i, p \in \mathcal{F}_2, j \in \mathcal{F}_3, m \in \mathcal{M}$).

- 1: initialize $t_1 = 0, t_2 = 0$, empty tabu list, $[y, z, w] = 0$.
- 2: **while** ($f_1(z, y, w) < \epsilon$)
- 3: $k = \arg \min_{k' \in \mathcal{F}_1, z_{k'}=0} c_k$. if $f_1(z, y, w) + c_k \leq \epsilon$, $z_k = 1$.
- 4: if $f_1(z, y, w) + c_k > \epsilon$ or $z = 1$, $p = \arg \min_{p' \in \mathcal{F}_2 \cup \mathcal{F}_3, y_{p'}=0, w_{p'}=0} c_k$.
- 5: if $f_1(z, y, w) + c_p \leq \epsilon$, open facility p ; else, break.
- 6: **end while**
- 7: Given $[z, y, w]$, get objective value $V_B = V(z, y, w)$ through Algorithm 1. $v_b = [z, y, w]$.
- 8: **while** ($t_1 < N_{t1}$)
- 9: compute $\mathcal{N}(z)$ with fixed $[y, w]$ and objective value through Algorithm 1, $p = \arg \min_{z' \in \mathcal{N}(z)} V(z', y, w)$
- 10: if $V(p, y, w) < V_B$; $V_B = V(p, y, w)$, $v_b = [p, y, w]$.
- 11: if $V(p, y, w) \geq V_B$, $p = \arg \min_{z' \in \mathcal{N}(z), z' \text{ is non-tabu}} V(z', y, w)$.
- 12: $z = p$, $t_1 = t_1 + 1$, update \mathcal{T}_1 .
- 13: **while** ($t_2 < N_{t2}$)
- 14: compute $\mathcal{N}(y, w)$ with fixed z and objective value through Algorithm 1, $q = \arg \min_{q' \in \mathcal{N}(y, w)} V(z, q)$
- 15: if $V(z, q) < V_B$; $V_B = V(z, q)$, $v_b = [z, q]$.
- 16: if $V(z, q) \geq V_B$, $q = \arg \min_{q' \in \mathcal{N}(y, w), q' \text{ is non-tabu}} V(z, q)$.
- 17: if any $q' \in \mathcal{N}(y, w)$ is tabu, deploy N_{div} rarely deployed facilities and clear \mathcal{T}_2 ; else, $[y, w] = q$, update \mathcal{T}_2 . $t_2 = t_2 + 1$.
- 18: **end while**
- 19: $t_2 = 0$
- 20: **end while**
- 21: get x and $V(z, y, w)$ with $v_b = [z, y, w]$ through Algorithm 1.

Output: $[z, y, w, x]$ and $V_B = V(z, y, w)$, $v_b = [z, y, w]$.

IV. SIMULATION RESULTS AND DISCUSSIONS

In our simulation, a 73GHz band is considered by exploiting the measurements in [13]. β is set as 0.046 [14]. For users, the SNR threshold, γ_t , is -10dB calculated from a 100Mbps outage rate with a 1GHz bandwidth [13]. The bandwidth of the backhaul link is also 1GHz and the noise power is $\sigma^2 = -74\text{dBm}$ [14] for mmW band. The access and backhaul outage probability are set 0.1. The subarea is set as $10\text{m} \times 10\text{m}$ and the density of users is $200/\text{km}^2$. The maximum number of relays for an SBS, N , is set as 2 and the backhaul capability of a BAN, N_b , is set as 5. θ is set to 0.5. For MTC, the results in [3] is used where an MA can support 600 machines with a latency requirement of 15 ms.

Fig. 2 shows a deployment instance which achieves the minimum objective $f_c(x)$, *best* coverage, in a $400\text{m} \times 400\text{m}$ area. In this area, 2000 machines are distributed uniformly. Here, the deployed BANs, SBSs, MAs and connection relationships are shown in the figure. Better coverage has been obtained through the proposed framework compared with conventional fiber-based and single-hop wireless backhaul models. In the result, 4 BANs have been deployed which provide backhaul links to 18 SBSs and 7 MAs. One potential site for BANs has not been used since the SBSs in the 2-hop wireless backhaul path of BAN 2 have covered the subareas around this potential site. Thus, through the utilization of sites without fiber connection, less BANs are needed compared with conventional models.

In Fig. 3, we show the Pareto optimal solutions and lower bounds obtained using the proposed algorithm for an $400\text{m} \times 400\text{m}$ area with 5 locations, 40 locations and 20 locations for

Algorithm 3 Algorithm to solve (P).

Input: $N_b, N_{max,L}, \Delta\epsilon, N_{t_1,max}, N_{t_2,max}$.

```

1: initialize  $\epsilon_0 = \sum_{k \in \mathcal{F}_1} c_k + \sum_{i \in \mathcal{F}_2} c_i + \sum_{j \in \mathcal{F}_3} c_j$ ,  $\mathcal{P} = \emptyset$ ,  $t = 0$ ,  $t_L = 0$ ,
    $\mathcal{B}_l = \emptyset$ .
2: while ( $\epsilon_t > \min_{k \in \mathcal{F}_1} c_k$ )
3:   initialize  $\lambda$ ,  $\mathcal{P}_r = \emptyset$ ,  $\mathcal{V}_r = \emptyset$ ,  $t_L = 0$ ,  $t_1 = 0$ ,  $t_2 = 0$ .
4:   while ( $t_L < N_{max,L}$ )
5:     solve  $(P_\lambda^t)$  through Algorithm 2,  $\mathcal{P}_r = \mathcal{P}_r \cup \{V_B\}$ ,  $\mathcal{V}_r = \mathcal{V}_r \cup \{v_b\}$ .
6:      $t_L = t_L + 1$ , update  $\lambda$ .
7:   end while
8:    $[z, y, w] = \arg \max_{[z', y', w'] \in \mathcal{V}_r} V(z', y', w')$ ,  $\mathcal{B}_l = \mathcal{B}_l \cup \{V(z, y, w)\}$ 
9:   while ( $t_1 < N_{t_1,max}$ )
10:    compute  $\mathcal{N}(z)$  with fixed  $[y, w]$ ,  $\mathcal{P}_e = \{[z', y, w] | z' \in \mathcal{N}(z), f_1(z', y, w) \in [\epsilon - \Delta\epsilon, \epsilon], [z', y, w] \text{ is nondominated}\}$ .
11:     $\mathcal{P} = \mathcal{P} \cup \mathcal{P}_e$ , delete dominated solutions in  $\mathcal{P}$ .
12:     $z = \{\text{best non-tabu solution in } \mathcal{P}_e\}$ ,  $t_1 = t_1 + 1$ , update  $\mathcal{T}_1$ .
13:    while ( $t_2 < N_{t_2,max}$ )
14:      compute  $\mathcal{N}(y, w)$  with fixed  $z$ ,  $\mathcal{P}_e = \{[z, y', w'] | [y', w'] \in \mathcal{N}(y, w), f_1(z, y', w') \in [\epsilon - \Delta\epsilon, \epsilon], [z, y', w'] \text{ is nondominated}\}$ .
15:       $\mathcal{P} = \mathcal{P} \cup \mathcal{P}_e$ , delete dominated solutions in  $\mathcal{P}$ .
16:       $[y, w] = \{\text{best non-tabu solution in } \mathcal{P}_e\}$ , update  $\mathcal{T}_2$ .  $t_2 = t_2 + 1$ .
17:    end while
18:     $t_2 = 0$ 
19:  end while
20:   $t_1 = 0$ ,  $t = t + 1$ , update  $\epsilon_t$ .
21: end while

```

Output: Nondominated solution set \mathcal{P} , a lower bound set \mathcal{B}_l .

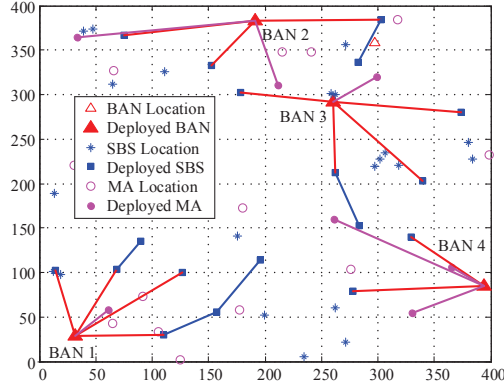


Fig. 2 Deployment instance when the number of potential locations for BANs, SBSs and MAs is 5, 40 and 20, respectively.

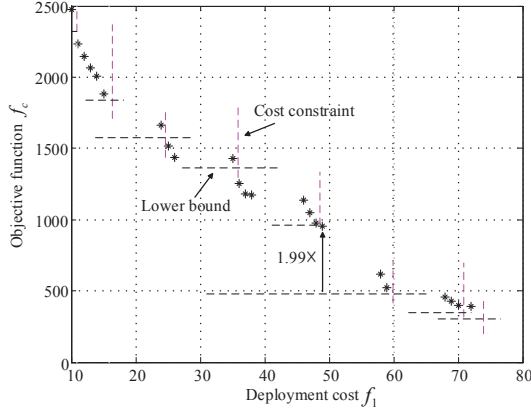


Fig. 3 Obtained Pareto optimal solutions and lower bounds.

BANs, SBSs and MAs, respectively. Here, the deployment costs

of an SBS and an MA is normalized to 1 and the deployment cost of a BAN is 10. Fig. 3 shows that the gap between the obtained solutions and the lower bound is quite tight. The maximum ratio of the objective value of the optimal solutions to the lower bound is 1.99. Clearly, the proposed algorithm is quite effective through the utilization of the structure of the problem.

V. CONCLUSIONS

In this paper, we have proposed a general deployment framework for next-generation cellular networks which incorporates HTC, MTC, wired backhaul links and multi-hop wireless backhaul links. A general multi-objective integer programming model has been proposed to determine the location of BANs, SBSs, MAs and their connections. To obtain the Pareto optimal solution set, we have proposed an adaptive heuristic algorithm by combining the weighted sum method, the ϵ -method, Lagrangian relaxation and tabu search jointly. The proposed algorithm can obtain both solutions and lower bounds which can be used to evaluate the quality of the obtained solutions. Simulation results have shown that the proposed framework can provide better deployment solutions than conventional fiber-based and single-hop wireless backhaul model. Furthermore, the objective values of obtained solutions are quite close to the lower bounds.

REFERENCES

- [1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5g be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [2] Senza Fili Consulting, "Small-cell backhaul: Industry trends and market overview," 2013.
- [3] K. Zhou, N. Nikaein, and R. Knopp, "Dynamic resource allocation for machine-type communications in lte/lte-a with contention-based access," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, April 2013, pp. 256–261.
- [4] E. Yaacoub and Z. Dawy, "On using relays with carrier aggregation for planning 5g networks supporting m2m traffic," in *Proc. of IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Larnaca, Cyprus, Oct 2014, pp. 124–129.
- [5] F. Richter and G. Fettweis, "Base station placement based on force fields," in *Proc. of IEEE 75th Vehicular Technology Conference (VTC Spring)*, Yokohama, Japan, May 2012, pp. 1–5.
- [6] P. Gonzalez-Brevis, J. Gondzio, Y. Fan, H. Poor, J. Thompson, I. Krikidis, and P.-J. Chung, "Base station location optimization for minimal energy consumption in wireless networks," in *Proc. of IEEE Vehicular Technology Conference (VTC Spring)*, Budapest, Hungary, May 2011, pp. 1–5.
- [7] H. Ghazizai, E. Yaacoub, M. Alouini, Z. Dawy, and A. Abu Dayya, "Optimized lte cell planning with varying spatial and temporal user densities," to appear in *IEEE Trans. Veh. Technol.*, no. 99, pp. 1–1, 2015.
- [8] H.-Y. Hsieh, S.-E. Wei, and C.-P. Chien, "Optimizing small cell deployment in arbitrary wireless networks with minimum service rate constraints," *IEEE Trans. Mob. Comput.*, vol. 13, no. 8, pp. 1801–1815, 2014.
- [9] W. Zhao, S. Wang, C. Wang, and X. Wu, "Cell planning for heterogeneous networks: An approximation algorithm," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, Toronto, Canada, April 2014, pp. 1087–1095.
- [10] S. Singh, M. N. Kulkarni, A. Ghosh, and J. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," in *Arxiv*, July 2014, <http://arxiv.org/abs/1407.5537>.
- [11] M. Islam, A. Sampath, A. Maharshi, O. Koymen, and N. Mandayam, "Wireless backhaul node placement for small cell networks," in *Proc. of the 48th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, USA, March 2014, pp. 1–6.
- [12] I. Park, D. Kim, and D. Har, "Mac achieving low latency and energy efficiency in hierarchical m2m networks with clustered nodes," *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1657–1661, March 2015.
- [13] A. Ghosh, T. Thomas, M. Cudak, R. Ratasuk, P. Moorut, F. Vook, T. Rapaport, G. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014.
- [14] M. Kulkarni, S. Singh, and J. Andrews, "Coverage and rate trends in dense urban mmwave cellular networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Austin, USA, Dec 2014, pp. 3809–3814.
- [15] A. F. Gabor and J.-K. C. W. van Ommeren, "A new approximation algorithm for the multilevel facility location problem," *Discrete Appl. Math.*, vol. 158, no. 5, pp. 453–460, Mar. 2010.
- [16] G. Kirlik and S. Sayin, "A new algorithm for generating all nondominated solutions of multiobjective discrete optimization problems," *European Journal of Operational Research*, vol. 232, no. 3, pp. 479 – 488, 2014.